Crystal structure of the ATPase domain of translation initiation factor 4A from *Saccharomyces cerevisiae* – the prototype of the DEAD box protein family

Jörg Benz[1]*, Hans Trachsel[2] and Ulrich Baumann[1]*

Background: Translation initiation factor 4A (eIF4A) is the prototype of the DEAD-box family of proteins. DEAD-box proteins are involved in a variety of cellular processes including splicing, ribosome biogenesis and RNA degradation. Energy from ATP hydrolysis is used to perform RNA unwinding during initiation of mRNA translation. The presence of eIF4A is required for the 43S preinitiation complex to bind to and scan the mRNA.

Results: We present here the crystal structure of the nucleotide-binding domain of eIF4A at 2.0 Å and the structures with bound adenosinediphosphate and adenosinetriphosphate at 2.2 Å and 2.4 Å resolution, respectively. The structure of the apo form of the enzyme has been determined by multiple isomorphous replacement. The ATPase domain contains a central seven-stranded β sheet flanked by nine α helices. Despite low sequence homology to the NTPase domains of RNA and DNA helicases, the three-dimensional fold of eIF4A is nearly identical to the DNA helicase PcrA of *Bacillus stearothermophilus* and to the RNA helicase NS3 of hepatitis C virus.

Conclusions: We have determined the crystal structure of the N-terminal domain of the eIF4A from yeast as the first structure of a member of the DEAD-box protein family. The complex of the protein with bound ADP and ATP offers insight into the mechanism of ATP hydrolysis and the transfer of energy to unwind RNA. The identical fold of the ATPase domain of the DNA helicase PcrA of *B. stearothermophilus* and the RNA helicase of hepatitis C virus suggests a common fold for all ATPase domains of DExx- and DEAD-box proteins.

Addresses: [1]Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, Germany and [2]Institut für Biochemie und Molekularbiologie, Universität Bern, Bühlstrasse 28, CH-3012 Bern, Switzerland.

*Corresponding authors.
E-mail: joerg.benz@ibc.unibe.ch
ulrich.baumann@ibc.unibe.ch

Key words: crystal, DEAD box, NTPase, translation initiation, X-ray

Received: 22 February 1999
Revisions requested: 18 March 1999
Revisions received: 19 April 1999
Accepted: 22 April 1999

Published: 1 June 1999

Structure June 1999, 7:671–679
http://biomednet.com/elecref/0969212600700671

 ISSN 0969-2126

Introduction

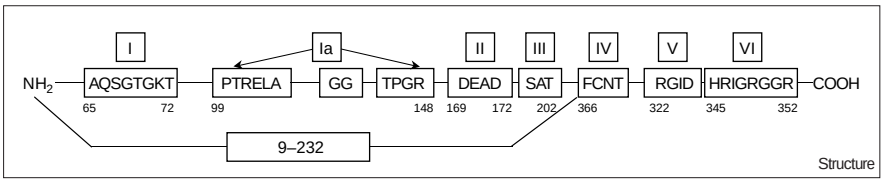
Initiation of translation in eukaryotes requires a complex biochemical pathway catalyzed by a large number of initiation factors (eIF) [1]. In the cap-dependent initiation pathway, translation initiation factor 4E (eIF4E) binds to the m7Gppp (cap) structure at the 5′ end of the mRNA and interacts with the scaffold protein eIF4G [2]. This large protein in turn binds to the initiation factors eIF4A [3], and probably also eIF4B and the poly(A)-binding protein, leading to mRNA circularization [4]. eIF4A together with eIF4B has RNA helicase or RNA unwinding activity and is thus believed to remove RNA secondary structure in the 5′ untranslated region of mRNA [5]. This creates a binding site on the mRNA for the 43S preinitiation complex, which recognizes the mRNA through interaction of ribosome-bound eIF3 with mRNA-bound eIF4G [6]. The ribosomal subunit then moves in the 5′ to 3′ direction from the cap-proximal region towards the initiator AUG codon in a process which was termed scanning [7] and which requires ATP hydrolysis, presumably for eIF4A-dependent RNA secondary structure melting [8].

eIF4A is the prototype of the DEAD-box protein family representing the core region common to all members, most of which possess additional variable extensions at their N and C termini [9]. Besides translation initiation, the DEAD-box family and related DExH- and DEAH-box proteins are involved in a variety of cellular processes including splicing, ribosome biogenesis and RNA degradation [10–12]. All DEAD-box proteins studied *in vitro* have been shown to be RNA-dependent ATPases, but the *in vivo* functions of most of the DEAD-box proteins have yet to be elucidated.

Sequence analysis and biochemical and genetic experiments with mutant eIF4A have revealed nine conserved sequence motifs (Figure 1). The ATPase A motif, also called Walker A motif (AxxGxGKT) [13], has been shown to be required for ATP binding, the ATPase B motif (DEAD) [9] is needed for ATP hydrolysis, and SAT and HRIGR motifs are important for helicase activity and protein–RNA interactions, respectively [14–17].

A kinetic model for the function of eIF4A has been suggested by Lorsch and Herschlag [18]. ATP hydrolysis is

Figure 1



Schematic representation of conserved regions in the sequences of RNA helicases of the DEAD-box family [12] and their location in eIF4A.

not required for RNA binding, but the state of the nucleotide is important for the conformation of the enzyme. Lorsch and Herschlag were able to show that eIF4A undergoes a cycle of ligand-induced conformational changes upon binding of ADP or ATP and RNA, which resulted in different sensitivity of eIF4A to proteolytic digestion by trypsin and chymotrypsin [19].

In the past few years, crystal structures of several RNA and DNA helicases have been determined [20–22]. In the NTPase domains of all of the structures at least the critical residues for ATP binding and hydrolysis are at the same position even if the rest of the protein is quite different. Here, we report the crystal structure of the nucleotide-binding domain of yeast eIF4A at 2.0 Å resolution and the crystal structures with bound ATP and ADP at 2.4 Å and 2.2 Å, respectively. The fold of the presented structure is very similar to those of the previously determined NTPase domains of DNA and RNA helicases and is nearly identical to the fold of PcrA, a DNA-helicase of *Bacillus stearothermophilus*, even though no sequence homology can be observed, suggesting a conserved fold for the nucleotide-binding domains in all helicases.

Results and discussion

Preparation and crystallization

So far, we have not succeeded in crystallizing full-length eIF4A. This may be explained by a polydisperse particle distribution, as revealed by light scattering experiments (data not shown). We therefore designed C-terminal deletion mutants of eIF4A based on the data of Lorsch and Herschlag [19]. They investigated the cleavage pattern generated by trypsin and chymotrypsin with mouse initiation factor 4A and were able to locate and identify stable domains of the protein. We expressed the N-terminal domain of yeast eIF4A including the Walker A and B motif and the SAT motif but lacking the RNA-binding region (Figure 1). As the sequence comparison of several DEAD-box proteins shows variable N termini, we also removed this part. The resulting fragment of eIF4A(9–232) was analyzed by light-scattering

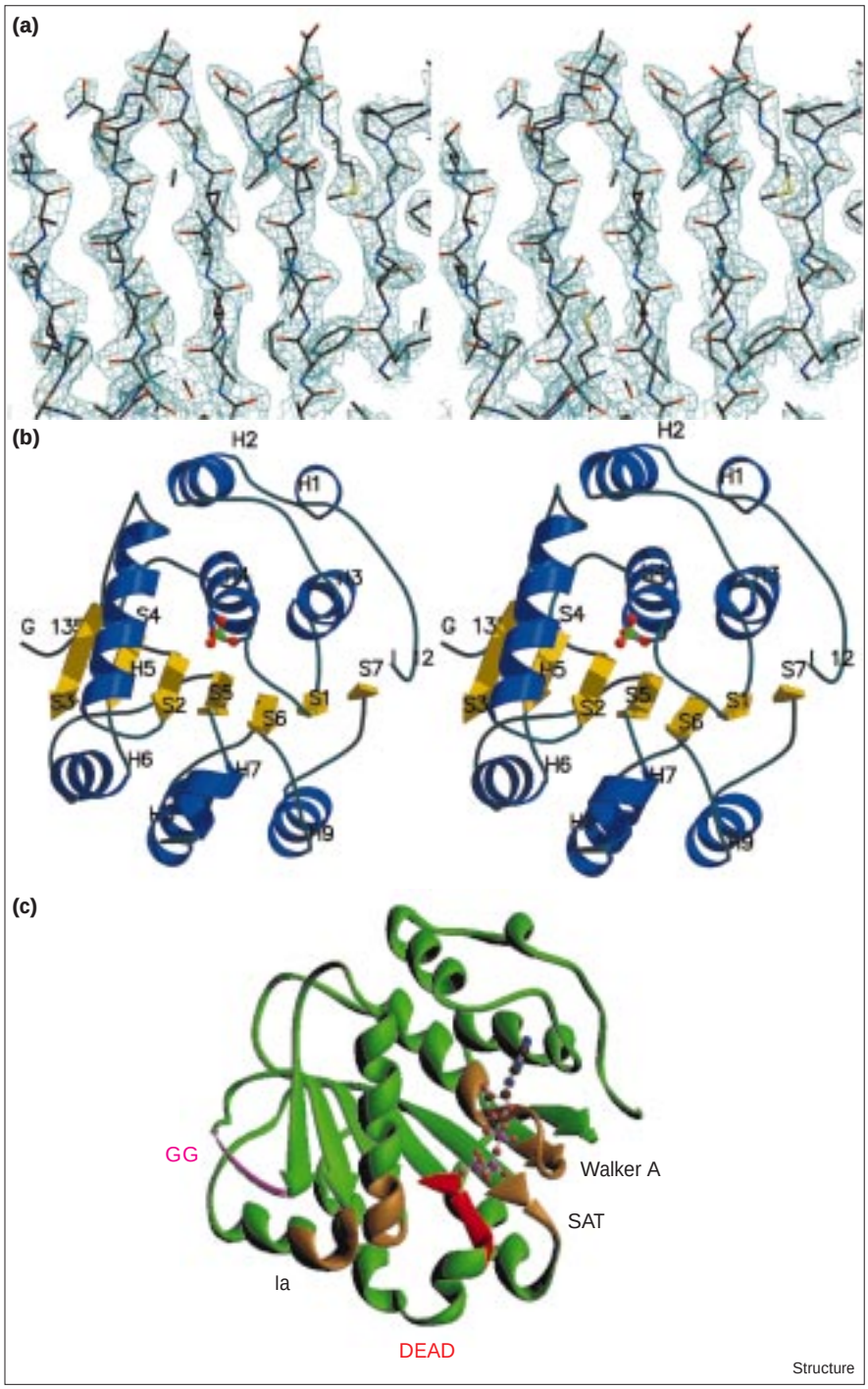
Table 1

Data collection and refinement statistics.

Data set	Native	K_3IrCl_6	$HgCl_2$	CH_3HgCl_2	ADP	ATP
Resolution (Å)	2.0	3.0	2.5	2.5	2.2	2.4
Reflections	14,989	4539	7586	7077	11,339	8773
Completeness (%)	99.6	97.9	96.3	89.9	97.4	98.8
R_{sym}*	9.0 (31.5)	4.5 (11.9)	6.0 (17.6)	7.4 (30.5)	6.4 (22.9)	9.6 (38)
Number of sites		1	1	1		
Phasing power†		1.70	1.54	1.80		
R_{cullis} (centric)		0.65	0.67	0.60		
Figure of merit	0.58 overall					
Refinement statistics						
Resolution (Å)	40.0–2.0			40.0–2.2	40.0–2.4	
Number of reflections	14,918				11,178	8708
Number of atoms	1587				1587	1587
Number of solvent molecules	49				42	43
Number of Mg ions					1	1
R_{fac}/R_{free} (%)	21.3/27.1				21.4/27.7	20.8/26.7
Rms deviation						
bonds	0.008				0.007	0.007
angles	1.38				1.25	1.28

*$R_{sym} = \Sigma\Sigma | (I(hkl;i) - <I(hkl>/\Sigma\Sigma I(hkl;i)$, where I(hkl;i) is the ith measurement of the reflection hkl; <I(hkl;i)> is the corresponding mean value. The values in parentheses represent the R_{sym} in the outermost resolution shell. †Phasing power = $f_H/<E>$, where f_H is the heavy-atom amplitude and <E> the mean lack of closure. The values were calculated using SHARP.

Figure 2



The overall fold of eIF4A. (a) Stereoview of a typical region of the solvent-flattened experimental electron-density map. The view is of the central β sheet, with carbon atoms in black, oxygen atoms in red and nitrogen atoms in blue. (b) Stereoview of a ribbon diagram of eIF4A(9–232). The yellow arrows represent β strands and the blue ribbons represent α helices. The figure was prepared with MOLSCRIPT [41] and rendered with Raster3D [42]. The α helices and β strands are labeled from H1 to H9 and from S1 to S7, according to the text. (c) Ribbon diagram of eIF4A(9–232). The conserved sequence elements of the DEAD-box family are labeled. The region with the GG motif is shown as a short stretch in magenta, the DEAD box is in red and the other conserved motifs are in orange.

experiments, which showed a monodisperse particle distribution in solution, and could then be crystallized. In contrast to the full-length protein, the truncated eIF4A(9–232) showed no activity in an *in vitro* eIF4A-dependent translation system [23] (data not shown). This is not surprising because eIF4A(9–232) lacks the RNA-binding motifs.

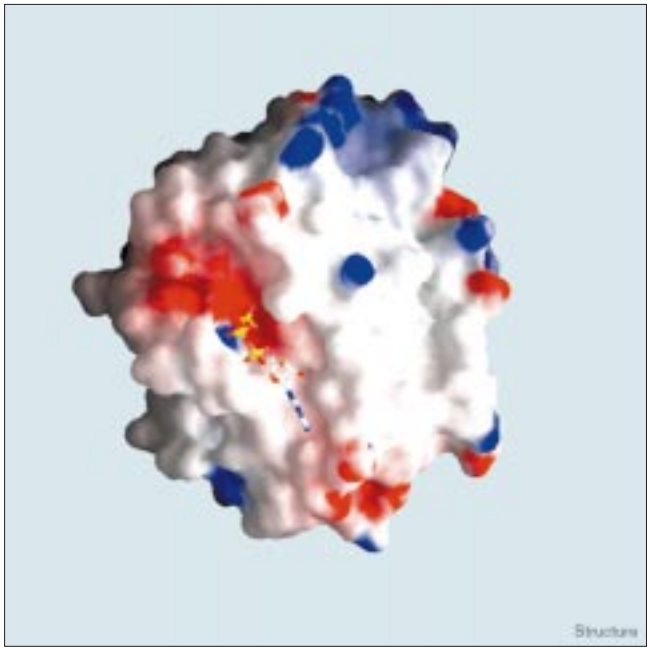
The structure of eIF4A(9–232)

The structure of eIF4A(9–232) was determined by the multiple isomorphous replacement method (Table 1). The protein crystallizes in space group $P2_12_12_1$ with one molecule in the asymmetric unit. The first solvent-flattened experimental electron density was very clear and most of the sidechains could be assigned (Figure 2a). No

electron density is seen for the first 15 residues, including the residues of the His tag and the cloning vector (MRGSHHHHHHGS), residues 9–11 of eIF4A(9–232), the stretch of amino acids 125–134 and the last nine amino acids at the C-terminal end of the protein. In total, 201 amino acids, 49 water molecules and one sulfate ion were included in the model of the apo form and the structure has been refined to a crystallographic R factor of 21.3% and R_{free} of 26.6% using data from 40 to 2.0 Å resolution (see the Materials and methods section and Table 1). The structure has good stereochemical parameters as estimated by the program PROCHECK [24], with 93% in the most-favoured region and 7% in additional-allowed regions in a Ramachandran plot.

eIF4A(9–232) has an αβα fold with overall dimensions of approximately $45 \times 45 \times 30$ Å. The protein contains a central, seven-stranded, twisted β sheet (S1(61–64), S2(92–95), S3(120–123), S4(141–144), S5(165–169), S6(195–200), S7(219–221)) flanked by nine α helices (H1(23–26), H2(30–38), H3(48–56), H4(71–82), H5(99–112), H6(146–154), H7(171–176), H8(180–189), H9(205–214)), with helices H1–H5 on one side and H6–H9 on the other side of the sheet (Figure 2b). H1 is formed by a one-turn 3_{10} α helix. All strands of the β sheet are parallel with a sheet topology of 7165243 characterized with the program PROMOTIF [25]. The N- and C-terminal ends are not visible in the electron density. The first defined element is the first loop, which begins with amino acid 12 and is tightly packed against the last visible segment, β strand S7. The N-terminal segment wraps around the molecule and leads into the first three helices. Helix H3 forms tight interactions with H4 and S1 and S7. Helix H4 is followed by a long loop and S2. The longest helix in the structure is H5 with 4 helical turns, which runs parallel to strand S3 and S4 but has nearly a perpendicular orientation to the plane of the rest of the β sheet. Regarding the orientation in Figure 2b, the upper part of the molecule is completed by S3. Between S3 and S4, no electron density is seen for nine amino acids. The lower part of the molecule is also characterized by three helices in parallel and helix H7 in perpendicular orientation with respect to the plane of the β sheet. The chain ends with strand S7 (Ile221). No density could be found for the last nine amino acids.

Figure 3



Electrostatic surface representation of eIF4A with bound ADP (stick representation), prepared with the program GRASP [43].

In the structure of the apo form of the enzyme the P loop or nucleotide-binding motif I coordinates a sulfate ion, which is stabilized by hydrogen bonds of the amide nitrogens from residues Gly68 to Thr72 (Figure 2b). A similar binding mode for a sulfate ion could be observed in the structure of the NS3 RNA helicase of hepatitis C virus [12], where the sulfate ion is also located at the binding site of the phosphate moiety of the nucleotide.

Location of the conserved sequence elements in the structure of eIF4A(9–232)

Up to five superfamilies of RNA helicases are known so far [11]. eIF4A belongs to superfamily II, which is characterized by nine conserved sequence elements [9]; six of them are retained in eIF4A(9–232) (Figure 1). With the exception of the GG motif, all conserved motifs are clustered in the structure of eIF4A (Figure 2c). In detail, the nucleotide-binding motif I or Walker A motif is located on the surface of the molecule in loop Gly68–Thr72 (between strand S1 and helix H4) and it helps to orient the phosphate moiety of the nucleotide in close distance to the DEAD box or motif II (Asp169–Asp172). Mutational analysis revealed that motif I is required for binding of the nucleotide and motif II for ATP hydrolysis [14–17]. Motif II is located at the end of strand S5 and in the connecting loop to helix H7. The residues of motif II form a hydrophilic, negatively charged pocket (Figure 3).

Extensive mutational studies have been carried out regarding the conserved sequence elements of the yeast and human eIF4A [14–17]. Exchange of Asp169 (Asp182 in human eIF4A) with either histidine or glutamate leads to a lethal phenotype in yeast. In human eIF4A, replacement of Asp182 by asparagine abolished the ATPase and helicase activity, the mutation Asp182 to glutamate reduced ATPase activity and no helicase acitivity could be measured [16]. This can be explained directly from the structure, where the negative charge of the aspartic acid is involved in interactions with the conserved Lys71 and

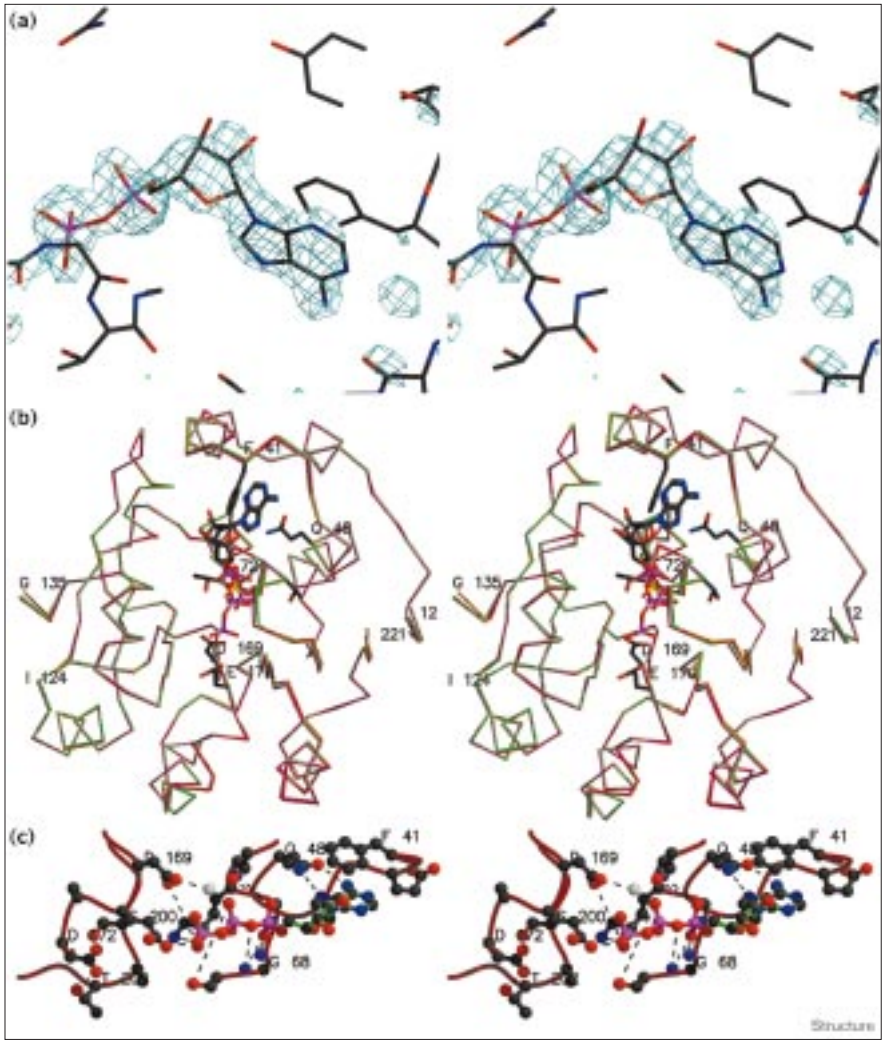
Thr72 and with a putative magnesium ion and water molecules. The distance between the first aspartic acid of the DEAD-box motif and the γ phosphate of the ATP is only 3.2 Å. Therefore, elongation of the sidechain results in steric hindrance.

An important role is played by Lys71, which points towards the Walker B motif and forms a salt bridge to Asp169 in the apo form. If the lysine sidechain is replaced by an uncharged amino acid like asparagine, binding of ATP is abolished. In the ADP and ATP complex a closer interaction with Glu170 of the DEAD-box motif is formed (Figure 4c). Asp172 is within hydrogen-bonding distance of Thr202 of the SAT motif, which is thought to be required for the helicase activity of eIF4A [17]. The residues Ser200–Thr202 are located in a loop connecting strand S6 and helix H9. A conformational switch was proposed for the interaction of the serine and threonine residue with the DEAD box because of the fact that mutations in both Ser200 and Thr202 abolished helicase activity [17] and a slight rearrangement of the hydrogen bond from Thr202 to Ser200 is possible.

The elements of motif Ia (Figure 1) are located on the surface in loops after β strands S2 (Pro99–Ala104) and S4 (Thr145 to Arg148). In this part of the molecule a cluster of positive charges is formed on the surface. Because of their close proximity to the SAT motif, it is possible that these residues are also involved in RNA binding and helicase activity. This suggestion is supported by comparison of the structure of eIF4A(9–232) with that of NS3 helicase of hepatitis C virus with bound oligonucleotide [22]. With regard to the conserved sequence elements, the first domain in NS3 harbours the ATPase function, as in eIF4A, and the second domain binds to RNA. Superposition of NS3 with eIF4A reveals that the DNA stretch is

Figure 4

Nucleotide binding by eIF4A(9–232). (a) Stereoview of the ADP-binding site. The F_o–F_c electron density of ADP was calculated before ADP was included in the model; the contour level is 2.9σ. (b) Stereoview of an overlay of the apo form of eIF4A(9–232) with the ADP and the ATP complex in a Cα representation. Apo eIF4A is colored in red, the ADP complex is in green and the ATP is in magenta. The red patch represents the electrostatic potential of the DEAD motif. (c) Stereoview of the nucleotide environment of eIF4A(9–232). The residues involved in nucleotide binding and ATP hydrolysis are labeled.



not only bound to the second domain of the protein but also located closely to loops in the first domain corresponding to motif Ia in eIF4A (9–232).

The GG motif, which is also important for the activity of eIF4A, is located in the disordered stretch of amino acids that connects β strands S3 and S4 and is therefore not seen in the electron density. Mutating the glycine residues of the GG motif to aspartic acid leads to a lethal phenotype in yeast. These results cannot be explained from the structure. A possible explanation might be that these residues, or this region of the molecule, are involved in binding to the initiation factor 4G. Binding of eIF4A to eIF4G is necessary for protein translation [1].

The structure of eIF4A(9–232) with bound ADP and ATP

The complexes with ADP and ATP were obtained by soaking the crystals of the apo form of the enzyme with a highly concentrated solution of the ligand. In the ADP and ATP structures, each ligand is bound in a stable way (Figures 4a–c). The ATPase activity of eIF4A is stimulated by RNA and therefore in eIF4A(9–232), which lacks the RNA-binding motif, ATP is not hydrolyzed and can be located in the crystal structure. In order to ensure that our crystals contained ATP and not hydrolyzed ADP, we undertook ATP-hydrolysis assays. In the absence of protein, a background Pi release corresponding to 2.5–5% of the total ATP content in the reaction mixture could be detected during a 15 h reaction time. In the crystallization and soaking buffer, no ATPase activity could be observed for eIF4A(9–232) and full-length eIF4A without RNA on a time scale of 15 h. In a buffer resembling conditions of the cytosol, 1 pmol eIF4A hydrolyzed 1–3 pmol ATP/hour and 1 pmol eIF4A(9–232) hydrolyzed 0.2–0.5 pmol ATP per hour. These results are in accord with the ones reported for full-length eIF4A [16]. As a further hint that we really observe an ATP complex we calculated a F_o(ATP)–F_o(ADP) map using phases from the refined apo model excluding the SO_4 ion. This map shows three prominent peaks: the highest one (8.8σ) is at the position of the γ phosphate of the ATP model. Two other peaks of height 5.6σ are at the positions of the putative Mg^{2+} coordinated to the γ phosphate of the ATP model and close to the position of Thr72, which in the ATP structure is slightly shifted together with its adjacent residues in comparison to the ADP complex. The most prominent negative features in the map are along the positions of the ADP atoms of the ADP complex. ADP has a higher binding affinity for eIF4A than ATP does [18] and the negative feature in the F_o(ATP)–F_o(ADP) map is in accordance with the observation that ADP occupies the binding site more often than ATP does. Nevertheless, the highest positive feature in the map is the γ phosphate; this position is not occupied by sulfate in the ADP complex or in the apo form of eIF4A.

The binding mode of the nucleotides is analogous to the one observed in the structure of the DExx helicase PcrA [20] or in other P-loop enzymes [26]. The nucleotide is bound almost perpendicularly to the β sheet of eIF4A and the ligand is fixed at two sites in the structure. The adenine base fits into a pocket built of residues conserved in all DEAD-box proteins. The adenine moiety makes hydrophobic van der Waals contacts to the conserved Phe41. The aromatic ring of the phenylalanine is flipped into a parallel orientation with regard to the adenine when compared with the apo form. The mainchain oxygen of Glu43 and the sidechain of Gln48, which is conserved throughout the DEAD-box proteins, form hydrogen bonds to N6 and N7 of the adenine. No interactions are found between the ribose and the protein. The phosphate moiety is positioned by the glycine-rich loop of motif I. Hydrogen bonds are formed between the amide nitrogens of residues Gly68–Thr72 and the phosphate groups. In an overlay of the ADP and ATP structure with the apo form, the sulfate ion is located halfway between the α and β phosphates. Presumably a magnesium cation is coordinated to the β phosphate in ADP, and in the ATP structure it bridges the β and γ phosphates. Water molecules and the conserved Thr72 provide further ligands for the cation. The putative magnesium ion also makes contacts with the sidechain of Asp169, which is thought to be involved in hydrolysis of the γ phosphate. The invariant Lys71 interacts by polar contacts with the β- and γ-phosphate groups of the bound nucleotide and is close in distance to Glu170 of the DEAD-box motif. Almost no differences can be observed upon binding of the nucleotide in a comparison of the wild-type structure and the two complex structures. Superimposing all protein atoms for the apo structure and the one with bound ADP or ATP leads to root mean square (rms) values of 0.4 Å and 0.59 Å, respectively (Figure 4b). The largest differences are observed in motif I. Upon binding of ADP or ATP, the nucleotide-binding loop shifts from the position it occupies in the apo form by about 0.5 Å.

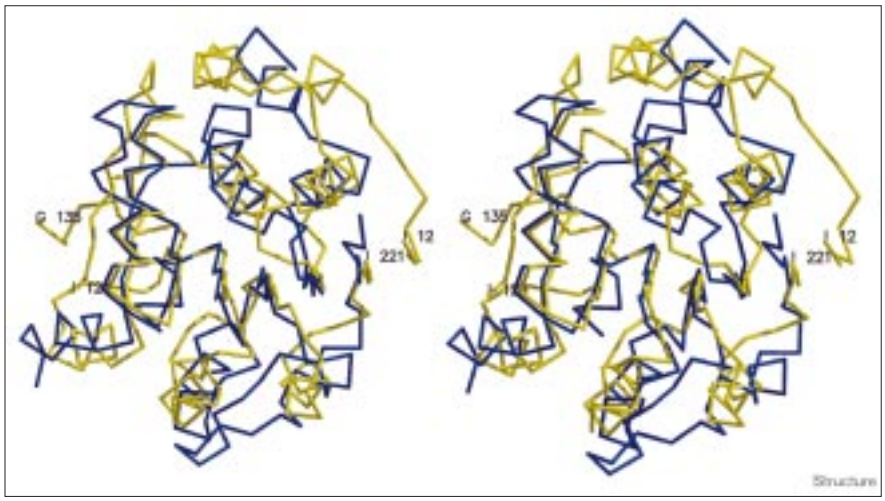
Only small changes can be observed for the other conserved motifs upon nucleotide binding, and the hydrogen-bonding pattern between Asp172 and Thr202 or Ser200 is also not changed, probably due to the lack of the RNA-binding domain. The hydrogen bond between the Oε of Asp172 and the hydroxyl group of Thr202 has a distance of 2.8 Å in the apo form and 3.0 Å in the form with bound nucleotide.

A possible mechanism for ATP hydrolysis

The structural similarities in the nucleotide-binding motif I and in the motif II to other nucleotide hydrolases suggest an identical or at least similar mechanism for hydrolysis of ATP [26]. Motif II (DEAD) is structurally conserved in those proteins, exhibiting a change in conformation upon ATP hydrolysis. This is then used to unwind

Figure 5

Stereoview of an overlay of eIF4A(9–232) and PcrA. eIF4A is shown in yellow and PcrA is in blue. The view is identical to that in Figure 2b. The figure was prepared with the program MOLSCRIPT.



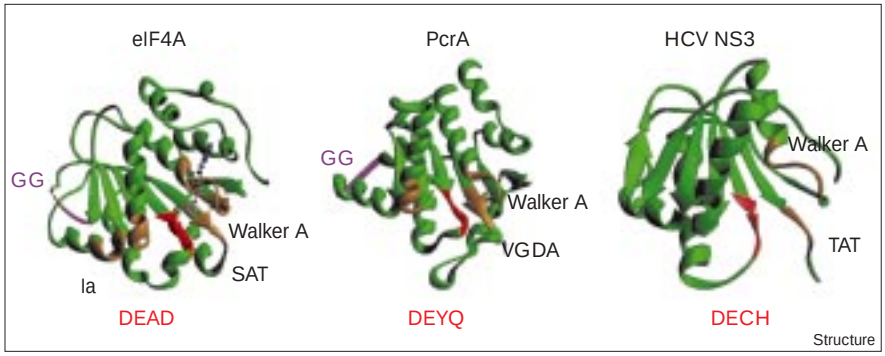
RNA in the DEAD-box family [18] or DNA in the DExx family [27]. In eIF4A the conformational change is probably a consequence of change in hydrogen-bonding pattern in the environment of the SAT motif. The hydrolysis of the γ phosphate requires a catalytic base, a function that can be fulfilled by aspartic acid 169 on the opposite site of the nucleotide-binding loop (Figure 4c). In the mechanism suggested here and for other nucleotide hydrolases [26] the catalytic base activates a water molecule and the cleavage of the bond is performed by an attack on the γ phosphate by the water molecule.

Structural comparison with other ATPases

We used the programs DEJAVU [28] and DALI [29] to search for a similar fold in the structure of eIF4A(9–232) and in nucleotide-binding domains of other proteins in the database . The similarity to the fold of the DExx DNA helicase PcrA of *B. stearothermophilus* [19] is striking (Figures 5 and 6). The two structures superimpose with a rms value of 2.5 Å for all Cα atoms of the nucleotide-binding domain. The major difference is the longer N terminus in eIF4A, with the additional first helix and an additional β strand (S3) resulting in a seven-stranded β sheet instead of a six-stranded sheet in the PcrA structure. In PcrA, strand S3 of eIF4A is replaced by a large insertion of 113 amino acids. Both structures are identical in the positions of the conserved motifs, suggesting an identical mechanism for transferring the energy of ATP hydrolysis to the RNA or DNA. The nucleotide-binding domains of NS3 of hepatitis C virus [21,22] and RecA [30] are also related in structure. In the hepatitis C virus RNA helicase NS3, the conserved elements can again be found at identical positions. The homologous regions begin with strand S2 of eIF4A and most of the secondary-structure elements are conserved. In the NS3 helicase, a β strand in the structure complements the undefined region in the electron density around the GG motif (Ile125–Gly134) of eIF4A(9–232). Additional parts in the eIF4A are helix H9 and strand S7. The conserved SAT motif in eIF4A(9–232) is therefore located in a loop connecting these two structural elements. In the NS3 helicase,

Figure 6

Ribbon plot of a side-by-side view of PcrA DNA helicase, eIF4A(9–232) and NS3 helicase. Molecules and conserved sequence elements are labeled. The conserved sequence elements are in orange, the DEAD-box motif is in red and the GG motif is in magenta.



the TAT motif is found in the hinge region between domains I and II (Figure 6).

A common structural theme for helicases has already been discussed on the basis of the crystal structures of PcrA DNA helicase, the NS3 RNA helicase and the Rep DNA helicase [27,31]. The highly similar structure of the nucleotide-binding domain of eIF4A supports this conclusion and suggests a similar fold for all nucleotide-binding domains of RNA and DNA helicases.

Biological implications

Protein translation is a very basic event during the life of a eukaryotic cell and control of mRNA translation plays an important role in regulation of gene expression. For the initiation of mRNA translation, several proteins are needed to form a preinitiation complex. The eukaryotic initiation factor 4A is part of the preinitiation complex, and if the gene is removed no mRNA translation can be performed. eIF4A represents the prototype of the DEAD-box proteins, a family of RNA helicases, and exhibits RNA-dependent ATPase activity and ATPase-dependent RNA-unwinding activity, leading to the conclusion that eIF4A is involved in removing RNA secondary structures, a function that is necessary for ribosomal scanning.

The structure of the ATPase domain of the prototype of DEAD-box proteins was determined by the multiple isomorphous replacement method. We were also able to characterize the complexes of eIF4A with ADP and ATP. The comparison of the two structures with that of the apo form of the enzyme gives us an insight into the mechanism of ATP hydrolysis in a DEAD-box protein. eIF4A shows high structural similarity to the ATPase domains of other helicases, namely, PcrA of *Bacillus stearothermophilus* and the NS3 helicase of hepatitis C virus. This suggests a similar or even identical fold for ATPase domains of RNA and DNA helicases with regard to the secondary-structure elements or the position of the conserved sequence elements involved in ATP hydrolysis.

Materials and methods

Cloning and protein preparation

The gene for yeast eIF4A was amplified by polymerase chain reaction (PCR) from yeast genomic DNA (Stratagene), digested with the restriction endonucleases *Bam*HI and *Pst*I and cloned into the same sites of the expression vector pQE–30 (Qiagen). The protein was expressed as an N-terminally His-tagged fusion protein at 37°C for 4 h in the *Escherichia coli* strain JM109 and purified using a Ni-NTA-agarose column (Qiagen). Light-scattering experiments were performed with DynaPro from Protein Solutions.

The ATPase domain of eIF4A (residues 9–232) was cloned by PCR amplification of the plasmid containing the eIF4A gene, digested with *Bam*HI and *Pst*I and cloned into the same sites of the expression vector pQE–30 (Qiagen, Hilden). In addition to amino acids 9–232 of eIF4A, the purified protein contains the sequence MRGSHHHHHHGS from the expression vector at the N terminus. The ATPase domain was expressed as an N-terminally His-tagged fusion protein at room temperature for 12 h in the *E. coli* strain JM109. Cells were lysed by sonication and the insoluble cell material was removed by centrifugation at 15,000 × g for 10 min. The supernatant was loaded onto a Ni-NTA-agarose column (Qiagen) and the protein was eluted with 120 mM imidazole. The protein preparation of eIF4A(9–232) was estimated by sodium dodecyl sulphate (SDS) gel electrophoresis to be more than 99% pure. Yields for eIF4A(9–232) were about 100 mg/l of culture. A gel-filtration run on a Superdex S75 (16/60) column (Pharmacia) resulted in a single peak of about 26 kDa, which corresponds to a monomer of eIF4A(9–232).

ATPase assay

The ATPase activity was measured in a reaction mixture containing both radioactively labeled [γ^{32}P]ATP and ATP. The measurements were performed either in 20 mM Tris-HCl, 100 mM KCl, 2 mM magnesium acetate or in the crystallization buffer. [γ^{32}P]ATP and ^{32}P were separated on thin layer chromatography sheets (PEI cellulose F, Merck) with 250 mM LiCl, 1 M formic acid as running solvent. The appropriate bands were cut out and evaluated in a scintillation counter.

Crystallization

Prior to crystallization the protein was dialysed against 5 mM Tris-HCl (pH 8.8) and concentrated to 20 mg/ml. Crystals were grown at 20°C in hanging drops by vapor diffusion from 20% PEG 8000, 0.01 M $MgCl_2$, 0.2 M $(NH_4)_2SO_4$ and 0.1 M MES (pH 4.8).

Data collection and phasing

All diffraction data were collected at 20°C on a R-axis IV image plate area detector system mounted on a Rigaku RU200 rotating-anode generator operating at 100 mA and 50 kV. Intensities were integrated, merged and scaled using the programs DENZO and SCALEPACK [32]. The crystals belong to the orthorhombic spacegroup $P2_12_12_1$, with unit-cell dimensions a = 38.9 Å, b = 59.3 Å and c = 92.1 Å. The crystals contain one molecule per asymmetric unit. The structure was solved by the multiple isomorphous replacement (MIR) method using two mercury, one gold and one iridium heavy-atom deriviative (Table 1). Crystals were soaked in the crystallisation solution containing 1 mM (mercury and gold) and 10 mM (iridium) of the heavy-atom compound for 4–16 h (Table 1). The heavy-atom site in the mercury derivative was located by inspection of the Harker sections of difference Patterson maps and the position determined and confirmed by the use of the program SHELXS [33]. The other sites were determined by difference Fourier calculations and confirmed by difference Pattersons. All sites were refined using the program SHARP [34]. A 3.0 Å solvent-flattened map assuming 48% solvent content was calculated with SOLOMON using the MIR phases [35].

Model building

An initial model was built into this map using the program O [36]. The phases were improved further by combination with phases derived from the model (SIGMAA from the CCP4 package [37]). Initially, the model was refined by simulated annealing followed by positional refinement (X-PLOR [38] and CNS [39]). The results of the refinement are summarized in Table 1. During the rounds of refinement, rebuilding of the model was performed with O. The crystallographic R_{free} [40] was monitored at each stage to prevent model bias using 10% of the reflections as test set. No density could be observed for the first 16 amino acids, including the His-tag, and for amino acids 125–134.

Nucleotide complexes with eIF4A

The ADP and ATP complexes were obtained by transferring crystals into crystallization solution containing 10 mM ADP or ATP (SIGMA). The crystals were soaked for 16 h before data collection.

Accession numbers

The coordinates have been deposited with the RCSB. The accession code for the apo form of the nucleotide-binding domain of eIF4A is 1qde.

References

1. Pain, V.M. (1996). Initiation of protein synthesis in eukaryotic cells. *Eur. J. Biochem.* 236, 747-771.
2. Hentze, M.W. (1997). eIF4G: a multipurpose ribosome adapter? *Science* 275, 500-501.
3. Imataka, H. & Sonenberg, N. (1997). Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol. Cell. Biol.* 17, 6940-6947.
4. Wells, S.E., Hillner, P.E., Vale, R.D. & Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* 2, 135-140.
5. Rozen, F., Edery, I., Meerovitch, K., Dever, T.E., Merrick, W.C. & Sonenberg, N. (1990). Bidirectional RNA helicase activity of eukaryotic translation initiation factors 4A and 4F. *Mol. Cell. Biol.* 10, 1134-1144.
6. Lamphear, B.J., Kirchweger, R., Skern, T. & Rhoads, R.E. (1995). Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation. *J. Biol. Chem.* 270, 21975-21983 .
7. Kozak, M. (1989). The scanning model for translation: an update. *J. Cell Biol.* 108, 229-241.
8. Jaramillo, M., Dever, T.E., Merrick, W.C. & Sonenberg, N. (1991). RNA unwinding in translation: assembly of helicase complex intermediates comprising eukaryotic initiation factors eIF-4F and eIF-4B. *Mol. Cell. Biol.* 11, 5992-5997.
9. Linder, P., Lasko, P.F., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J. & Slonimsky, P.P. (1989). Birth of the DEAD-box. *Nature* 337, 121-122.
10. Schmid, S.R. & Linder, P. (1992). D-E-A-D protein familiy of putative RNA helicases. *Mol. Microbiol.* 6, 283-292.
11. Gorbalenya, A.E. & Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure–function relationships. *Curr. Opin. Struct. Biol.* 3, 419-429.
12. Lüking, A., Stahl, U. & Schmidt, U. (1998). The protein family of RNA helicases. *Crit. Rev. Biochem. Mol. Biol.* 33, 259-296.
13. Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945-956.
14. Pause, A. & Sonenberg, N. (1993). Helicases and RNA unwinding in translation. *Curr. Opin. Struct. Biol.* 3, 953-959.
15. Pause, A., Méthot, N. & Sonenberg, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol. Cell. Biol.* 13, 6789-6798.
16. Pause, A. & Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J. 11*, 2643-2654.
17. Schmid, S.R. & Linder, P. (1991). Translation initiation factor 4A from Saccharomyces cerevisiae: Analysis of residues conserved in the D-E-A-D family of RNA helicases. *Mol. Cell. Biol.* 11, 3463-3471.
18. Lorsch, J.R. & Herschlag, D. (1998). The DEAD box protein eIF4A. I. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* 37, 2180-2193.
19. Lorsch, J.R. & Herschlag, D. (1998). The DEAD box protein eIF4A. II. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry* 37, 2194-2206.
20. Subramanya, H.S., Bird, L.E., Brannigan, J.A. & Wigley, D.B. (1996). Crystal structure of a Dexx box DNA helicase. *Nature* 384, 379-383.
21. Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A.D., Le, H.V. & Weber, P.C. (1997). The structure of the hepatitis C virus RNA helicase domain. *Nat. Struct. Biol.* 4, 463-467.
22. Kim, J.L., *et al*, & Carron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* 6, 89-100.
23. Altmann, M., Edery, I., Sonenberg, N. & Trachsel, H. (1985). Purification and characterisation of protein synthesis initiation factor eIF-4E from the yeast *Saccharomyces cerevisiae*. *Biochemistry* 24, 6085-6089.
24. Laskowski, R.A., McArthur, M.W., Moss, D.S. & Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291.
25. Hutchinson, E.G. & Thornton, J.M. (1996). PROMOTIF - a program to identify and analyze structural motifs in proteins. *Prot. Sci.* 5, 212-220.
26. Smith, C. & Rayment, I. (1996). Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys. J.* 70, 1590-1602.
27. Bird, L.E., Subramanya, H.S. & Wigley, D.B. (1998). Helicases: a unifying structural theme? *Curr. Opin. Struct. Biol.* 8, 14-18.
28. Kleywegt, G.J. & Jones, T.A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* 277, 525-546.
29. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123-131.
30. Story, R.M. & Steitz, T.A. (1992). The structure of the *E. coli* RecA protein monomer and polymer. *Nature* 355, 318-325.
31. Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M. & Waksman, G. (1997). Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 90, 635-647.
32. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307-326.
33. Sheldrick, G.M. (1997). Patterson superposition and ab initio phasing. *Methods Enzymol.* 276, 628-641.
34. De La Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiple wavelength anomalous diffraction methods. *Methods Enzymol.* 276, 472-494.
35. Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F_1ATPase. *Acta Crystallogr. D* 52, 30-42.
36. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110-119.
37. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50, 760-763.
38. Brünger, A.T. (1992). X-PLOR Version 3.1. *A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
39. Brünger, A.T., *et al*., & Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905-921.
40. Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-474.
41. Kraulis, P.J. (1991). MOLSCRIPT: a program for producing both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.
42. Merrit, E.A. & Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* 277, 505-524.
43. Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP - graphical representation and analysis of surface properties. *Biophys. J.* 64, A166.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.